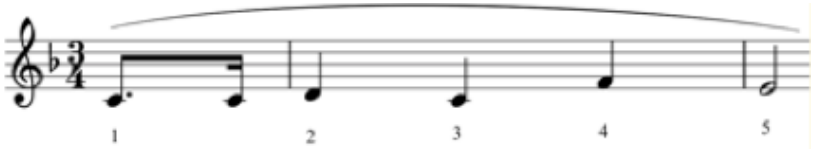
The evaluation of singing voice accuracy: Are we good judges even if we are not musicians?

SMPC
08/08/2013

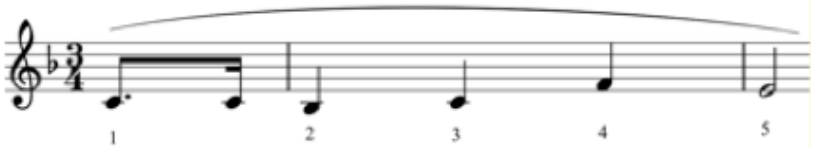
Larrouy-Maestri P., Roig-Sanchis V., Morsomme D.

Department of Psychology – University of Liège - Belgium

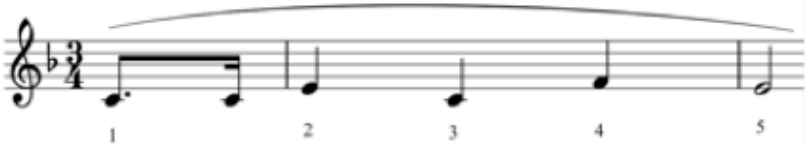
What is “in tune” ?



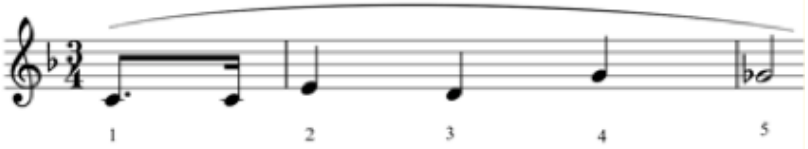
Accurate version



Contour error at the note 2



Interval errors of 200 cents between notes 1-2 and 2-3



Tonality modulation at the note 2

How to evaluate the vocal accuracy ?

Overview

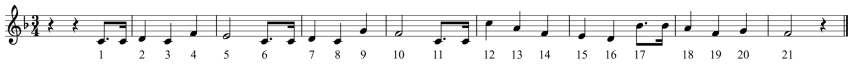
- Self-evaluation
  - Asking the singer himself
    - **Difficult** (Cuddy et al., 2005; Sloboda et al., 2005; Wise & Sloboda, 2008)
    - **Under-estimation** (Pfordresher & Brown, 2007)
- Music experts
  - Grids, scales, global score
    - e.g. Alcock et al., 2000a, 2000b; Hébert et al., 2003; Lévêque et al., 2012; Racette et al., 2006; Schön et al., 2004; Wise & Sloboda, 2008
- Computer-assisted method
  - Acoustic analyses, extraction F0, computation of errors
    - e.g. Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007; Lévêque et al., 2009; Pfordresher & Brown, 2007; Pfordresher et al., 2010

Are music experts good judges?

Larrouy-Maestri, Lévêque, Schön, Giovanni, & Morsomme, D. (2013)

Acoustic analyses

166 sung performances



http://sldr.org/sldr000774/en

18 Musicians

Procedure described in Larrouy-Maestri & Morsomme (in press)

-Estimation of the f0 for each note

-Computation of errors

- Number of contour errors
- Mean interval deviation
- Number of modulations

1----2----3----4----5----6----7----8----9

very inaccurate very accurate

Are music experts good judges?

Larrouy-Maestri, Lévèque, Schön, Giovanni, & Morsomme, D. (2013)

- Conclusions
  - High correlation between the raters ($r = .77$, $p < .01$)
    - Same definition of vocal accuracy
  - High correlation between the two methods ($r = .87$, $p < .001$)
    - Objectivity of the judges
    - 3 judges are sufficient to keep a correlation of .85
  - 81% of the variance of the judges is explained
    - Pitch interval deviation ($\beta = 0.51$; $p < .001$)
    - Number of tonality modulations ($\beta = 0.45$; $p < .001$)
- Answer...

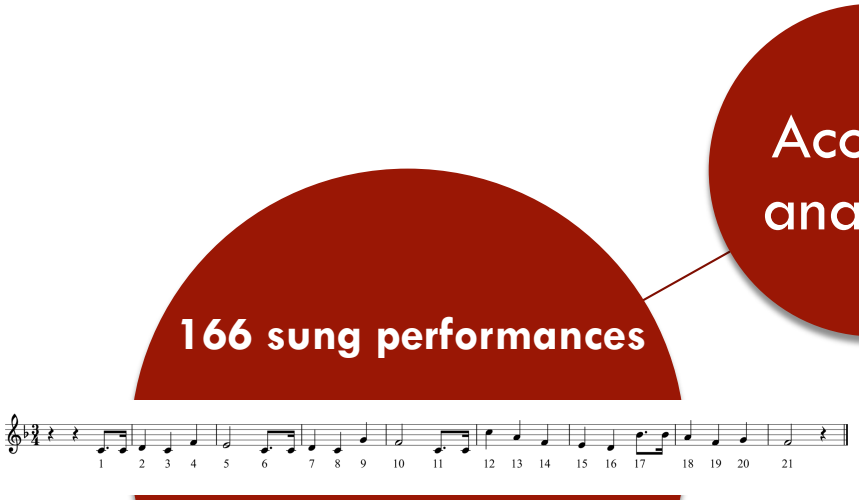
But …

Effects of music expertise	We are all experts of our culture (Bigand & Delbé, 2010)
• **Discrimination** Micheyl et al., 2006; Tervaniemi et al., 2005 • **Pitch perception** Hutchins & Peretz, 2012 • **Interval size estimation** Russo & Thompson, 2005 • **Complex tones** Hutchins et al., 2012; Vurma et al., 2010; Zarate et al., 2012 • **Error detection** Fujiroka et al., 2004; Hutchins et al., 2012 ; Warrier & Zatorre, 2002 • **Melodic recognition** Orsmond & Miller, 1999	• **Early development of melodic perception** Chang & Trehub, 1977; Ferland & Mendelson, 1989; Plantinga & Trainor, 2005; Stalinski et al., 2008 • **Enculturation** Miyamoto, 2007; Stalinski & Schellenberg, 2012; Trainor, 2005 ; Trainor et al., 2012 • **Implicit learning** Jonaitis & Saffran, 2009; Loui et al., 2010; Saffran et al., 1999; Schön et al., 2008 • **Melodic expectations** Marmel et al., 2008 • **Ability to sing in tune** Dalla Bella et al., 2007

Are we good judges ?

- Methods
- Results
- Conclusions

Methods



Procedure described in Larrouy-Maestri & Morsomme (in press)

-Estimation of the f0 for each note

-Computation of errors

- Number of contour errors
- Mean interval deviation
- Number of modulations

1----2----3----4----5----6----7----8----9

very inaccurate very accurate

Methods

- Participants
  - Paired in age, gender, sociocultural background
  - Non musicians
    - Less than 2 years of musical training
    - Occasional listeners
  - Audio and MBEA OK
- Procedure
  - 166 x “Happy birthday”
    - 5 lists in random order
  - Global pitch accuracy

1----2----3----4----5----6----7----8----9
very inaccurate very accurate

  - Two times (test and retest with 8-15 days in between)

Are we good judges ?

- Methods
- Results
- Conclusions

Results and discussion

Judges reliability

- Intra-judge reliability
  - Mean Spearman correlation T1/T2: .66 (SD = .06)

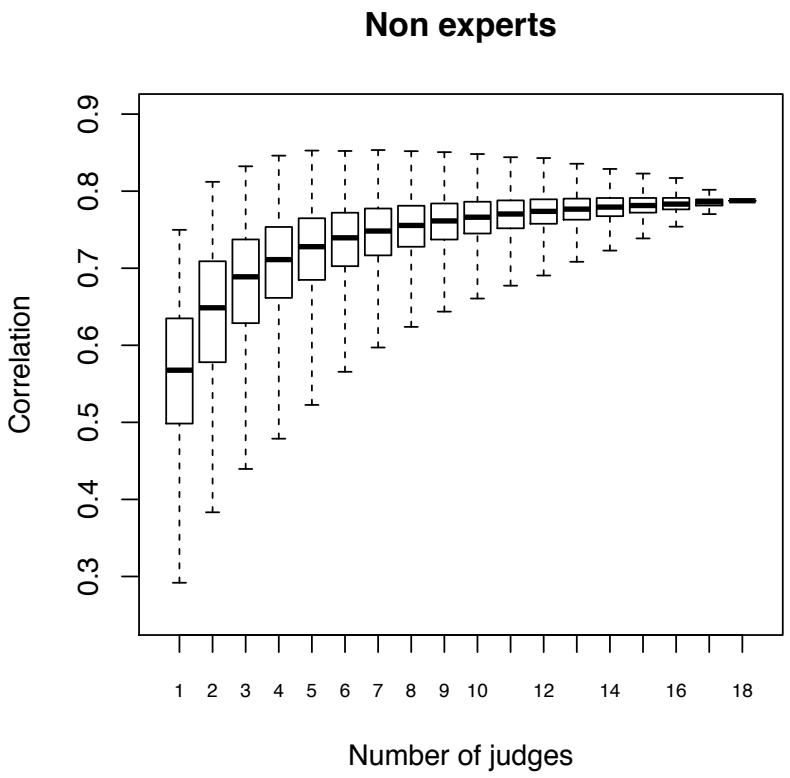
    ➔ **OK**
- Inter-judge reliability
  - Intra-class correlation coefficient: .89; $p < .01$

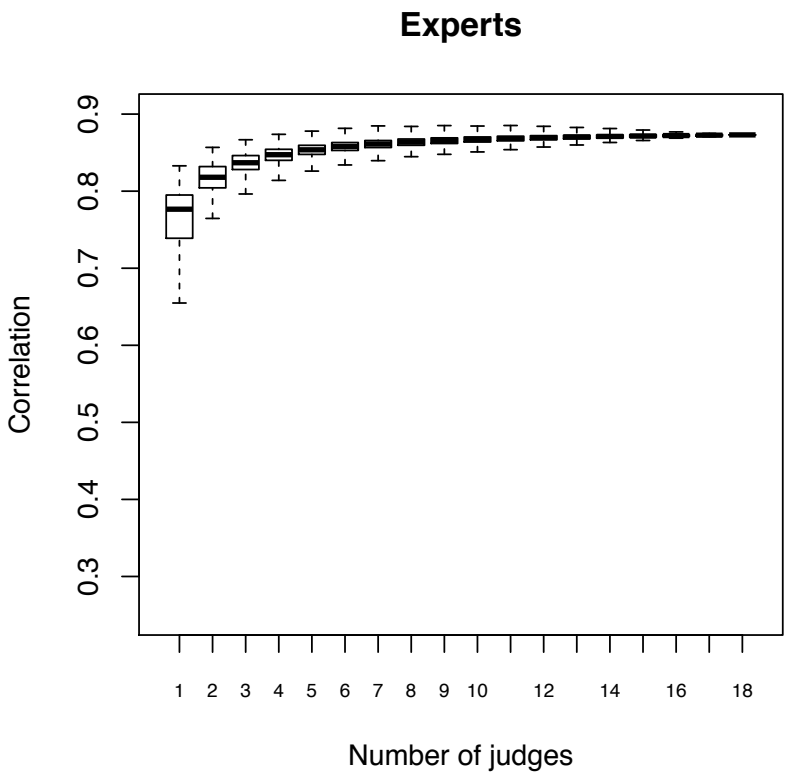
    ➔ **OK**
- Correlation with computer-assisted method
  - $r(166) = .81$; $p < .01$

    ➔ **Higher scores for accurate performances: OK**

Results and discussion

Correlation with decreased number of judges





Results and discussion

Comparison Non experts / Experts

- Comparison mean ratings
  - $r(166) = .84$; $p < .01$
  - *U Mann Whitney* $p < .01$
  - Non experts are more severe than experts
- Criteria

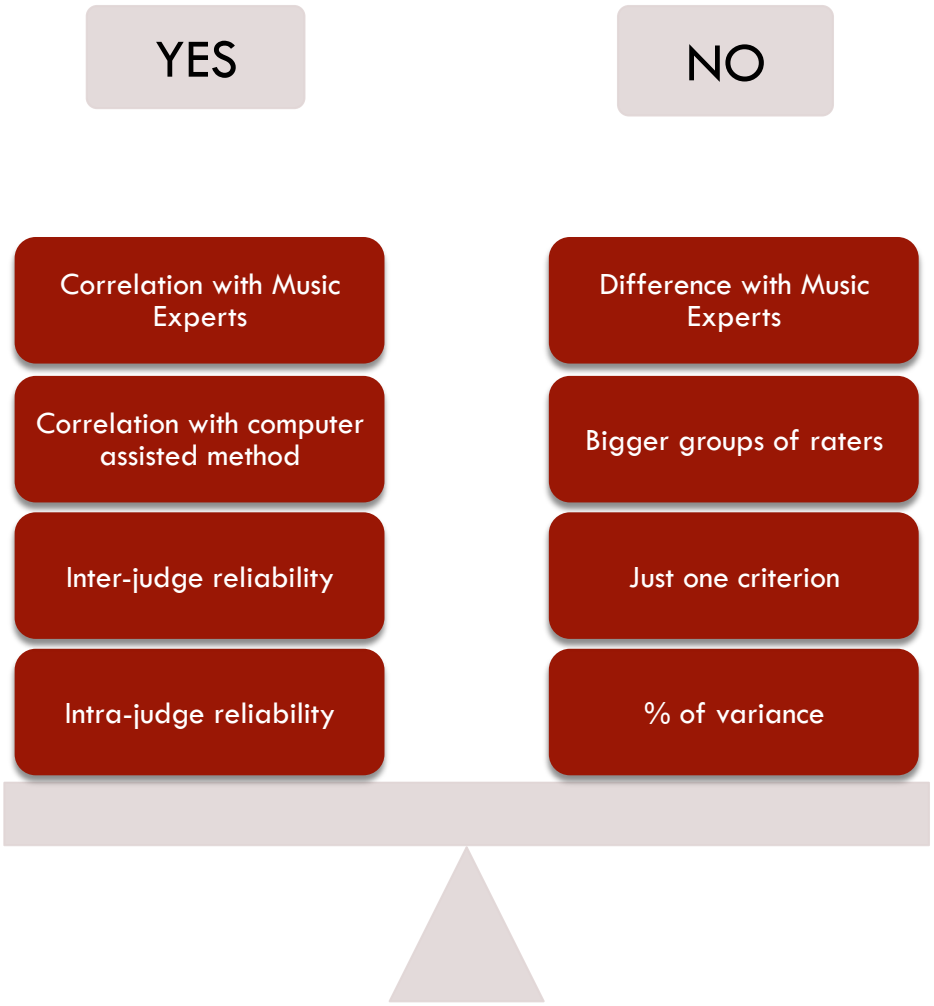
	Non experts	Experts
Model	$F(3,165) = 104.44$; $p < .01$	$F(3,165) = 231.51$; $p < .01$
% variance	66%	81%
Criteria	Interval deviation	Interval deviation Tonality modulations

Are we good judges ?

- Methods
- Results
- Conclusions

Conclusions

Are we good judges?



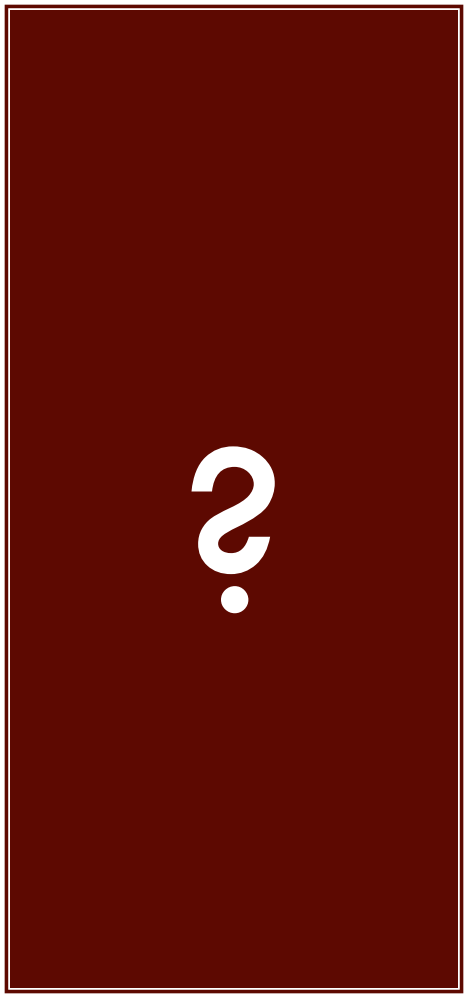
Conclusions

Are we good judges?

- Reliability and objectivity
  - Similar definition of vocal accuracy
  - Same strategy each time
  - Relationships with music experts' ratings and objective method
- However
  - More severe
  - Just sensitive to the precision of the intervals
  - 34% of the variance unexplained
- Perspectives
  - Other criteria which influence the rating of non musicians
  - Perception and tolerance of music experts and non experts

➔ Why?

The evaluation of singing voice accuracy



THANK YOU FOR YOUR ATTENTION

References

- Alcock, K. J., Passingham, R. E., Watkins, K., & Vargha-Khadem, F. (2000a). Pitch and timing abilities in inherited speech and language impairment. *Brain and Language, 75*(1), 34-46.
- Alcock, K. J., Wade, D., Anslow, P., & Passingham, R. E. (2000b). Pitch and timing abilities in adult left-hemisphere-dysphasic and right-hemisphere-damaged subjects. *Brain and Language, 75*(1), 47-65.
- Chang, H. W., & Trehub, S. E. (1977). Auditory processing of relational information by Young infants. *Journal of Experimental Child Psychology, 24*, 324-331.
- Cuddy, L. L., Balkwill, L. L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare: a study of "tone deafness" among university students. *Annals of the New York Academy of Sciences, 1060*, 311-324.
- Dalla Bella, S., & Berkowska, M. (2009). Singing Proficiency in the Majority. Annals of the New York Academy of Sciences, 1169(1), 99-107.
- Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. The Journal of the Acoustical Society of America, 121(2), 1182-1189.
- Ferland, M. B., & Mendelson, M. J. (1989). Infants' categorization of melodic contour. *Infant Behaviour Development, 12*, 341-355.
- Fujioka, T., Trainor L. J., Ross, B., Kakigi R., Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience. 16*, 1010-1021.
- Hébert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain, 126*(8), 1838-1850.

References

- Hutchins, S., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. Journal of Experimental Psychology: General, 141, 76–97.
- Hutchins, S., Roquet, C., & Peretz, I. (2012). The Vocal Generosity Effect: How Bad Can Your Singing Be? Music Perception, 30(2), 147-159.
- Jonaitis, E. M. M., & Saffran, J. R. (2009). Learning harmony: The role of serial statistics. *Cognitive Science, 33*, 951-968.
- Larrouy-Maestri, P., & Morsomme, D. (in press). Criteria and tools for objectively analysing the vocal accuracy of a popular song. *Logopedics Phoniatrics, Vocology*.
- Larrouy-Maestri, P., Lévêque, Y., Schön, D., Giovanni, A., & Morsomme, D. (2013). The Evaluation of Singing Voice Accuracy: A Comparison Between Subjective and Objective Methods. Journal of voice, 27(2), 259.
- Lévêque, Y., Amy de la Bretèque, B., Giovanni, A., & Schön, D. (2009). Les défauts de justesse de la voix chantée: compétences et déficits tonaux. *Revue de Laryngologie Otologie Rhinologie, 130*(1), 23-38.
- Lévêque, Y., Giovanni, A., & Schön, D. (2012). Pitch-matching in poor singers : Human model advantage. *Journal of voice, 26*(3), 293-298.
- Loui, P., Wessel, D. L., & Kam, C. L. H. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception, 27*(5), 377-388.
- Marmel, F., Tillmann, B., & Dowling, W. J. (2008). Tonal expectations influence pitch perception. *Perception & Psychophysics, 70*(5), 841-852.
- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*(1-2), 36-47.

References

- Miyamoto, K. A. (2007). Musical characteristics of preschool-age students : A review of literature. *National Association for Music Education, 26*, 26-40.
- Orsmond, G. I., & Miller, L. K. (1999). Cognitive, musical and environmental correlates of early music instruction. *Psychology of Music, 27*, 18-37.
- Plantinga, J., & Trainor, L. (2005). Memory for melody: infants use a relative pitch code. *Cognition, 98*(1), 1-11.
- Pfordresher, P. Q., & Brown, S. (2007). Poor-Pitch Singing in the Absence of "Tone Deafness". Music Perception, 25(2), 95-115.
- Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. The Journal of the Acoustical Society of America, 128(4), 2182-2190.
- Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: sing along! *Brain, 129*, 2571-2584.
- Russo, F. A., & Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics, 67*(4), 559-568.
- Saffran, J., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27-52.
- Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition, 106*(2), 975-983.
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*(3), 341-349.

References

- Sloboda, J. A., Wise, K. J. & Peretz, I. (2005). Quantifying tone deafness in the general population. *Annals of the New York Academy of Sciences, 1060*, 255–261.
- Stalinski, S. M., & Schellenberg, E. G. (2012). Music cognition: A developmental perspective. *Topics in Cognitive Science, 4*(4), 485-497.
- Stalinski, S. M., Schellenberg, E. G., & Trehub, S. E. (2008). Developmental changes in the perception of pitch contour. Distinguishing up from down. *Journal of the Acoustical Society of America, 124*, 1759-1763.
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch-discrimination accuracy in musicians vs. nonmusicians - an event-related potential and behavioral study. *Experimental Brain Research, 161*, 1–10.
- Trainor, L. J. (2005). Are there critical periods for musical development ? *Developmental Psychobiology, 46* (3), 262-278.
- Trainor, L. J., Marie, C., Gerry, D., Whiskin, E., & Unrau, A. (2012). Becoming musically enculturated: effects of music classes for infants on brain and behavior. *Annals of the New York Academy of Sciences, 1252*, 25-36.
- Vurma, A., Raju, M., & Kuuda, A. (2010). Does timbre affect pitch?: Estimations by musicians and non-musicians. *Psychology of Music, 39*(3), 291-306.
- Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. Perception & Psychophysics, 64(2), 198-207.
- Wise, K. J., & Sloboda, J. A. (2008). Establishing an empirical profile of self-defined "tone deafness": Perception, singing performance and self-assessment. *Musicae Scientiae, 12*(1), 3-26.